

1 1/2 / 149  
3/14/02

Attorney Docket No. 1359.1063

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Nobuyuki WASHIO

Application No.:

Group Art Unit:

Filed: February 26, 2002

Examiner:

For: SOUND SIGNAL RECOGNITION SYSTEM AND SOUND SIGNAL RECOGNITION METHOD, AND DIAGLOG CONTROL SYSTEM AND DIAGLOG CONTROL METHOD USING SOUND SIGNAL RECOGNITION SYSTEM



**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN  
APPLICATION IN ACCORDANCE  
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

Japanese Patent Application No. 2001-362996


Filed: November 28, 2001

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing date(s) as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: February 26, 2002

By:   
H. J. Staas  
Registration No. 22,010

700 11th Street, N.W., Ste. 500  
Washington, D.C. 20001  
(202) 434-1500

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日  
Date of Application:

2001年11月28日

出 願 番 号  
Application Number:

特願2001-362996

[ST.10/C]:

[JP2001-362996]

出 願 人  
Applicant(s):

富士通株式会社

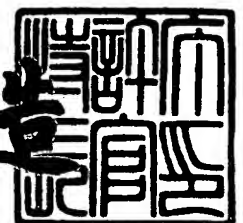


CERTIFIED COPY OF  
PRIORITY DOCUMENT

2002年 2月 1日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出証番号 出証特2002-3002732

【書類名】 特許願

【整理番号】 0195269

【提出日】 平成13年11月28日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/00

【発明の名称】 音信号認識システムおよび音信号認識方法並びに当該音信号認識システムを用いた対話制御システムおよび対話制御方法

【請求項の数】 11

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 鷲尾 信之

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 110000040

【氏名又は名称】 特許業務法人池内・佐藤アンドパートナーズ

【代表者】 池内 寛幸

【電話番号】 06-6135-6051

【手数料の表示】

【予納台帳番号】 139757

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0115801

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音信号認識システムおよび音信号認識方法並びに当該音信号認識システムを用いた対話制御システムおよび対話制御方法

【特許請求の範囲】

【請求項 1】 話音信号区間または D T M F 信号区間のいずれか一方または双方を含む音信号を入力する音信号入力部と、

話音信号モデルと、 D T M F 信号モデルを備え、前記音信号入力部から入力された音信号の照合処理において、前記話音信号モデルと前記 D T M F 信号モデルの双方を併せて参照に用いることにより照合処理を行なう照合部と、

言語モデルを備え、前記照合部の照合結果と前記言語モデルを用いて音信号の認識を行なう音信号認識部を備え、

前記話音信号区間または前記 D T M F 信号区間のいずれか一方または双方を含む音信号に対する音信号認識処理を実行することを特徴とする音信号認識システム。

【請求項 2】 前記音信号認識部が、

認識単位となる一区切りの音信号区間ごとに、前記照合部における前記話音信号モデルを用いた照合結果と前記 D T M F 信号モデルを用いた照合結果とを比較して照合結果の良い方を選択し、前記音信号区間ごとに選択された音信号認識結果をつないで前記入力音信号の全区間に対する音信号認識結果として統合する統合部を備えた請求項 1 に記載の音信号認識システム。

【請求項 3】 前記言語モデルが、 D T M F 信号を音信号認識語彙として含むことが可能な請求項 1 または 2 に記載の音信号認識システム

【請求項 4】 前記音信号入力部を介して音信号入力を行なう利用者に対して、特定の語彙について、話音による音信号入力とするか D T M F 信号による音信号入力とするかをガイダンスするガイダンス部を備えた請求項 1 から 3 のいずれかに記載の音信号認識システム。

【請求項 5】 前記統合部が、所定の条件により、特定の語彙について、話音により入力された音信号の誤認識率が高いことを検出した場合、前記ガイダンス部に対して、当該特定語彙について、 D T M F 信号により音信号を再入力する

ように求めるガイダンスを出力する指示情報を通知する請求項4に記載の音信号認識システム。

【請求項6】 前記統合部が、話音による音信号に対する照合結果における誤認識率と、DTMF信号による音信号に対する照合結果における誤認識率を予測、保持し、いずれか一方の誤認識率が所定値よりも高くなった場合、前記ガイダンス部に対して、他方の音信号による入力を促すガイダンスを表示する指示情報を通知する請求項4に記載の音信号認識システム。

【請求項7】 前記ガイダンス部が、DTMF信号と語彙の対応関係をあらかじめ利用者に通知する機能を備えた請求項4に記載の音信号認識システム。

【請求項8】 請求項1から7のいずれかに記載の音信号認識システムを含み、前記音信号認識システムによる音信号認識結果に基づいて利用者との対話の流れを制御する対話制御システム。

【請求項9】 話音信号区間またはDTMF信号区間のいずれか一方または双方を含む音信号を入力し、

話音信号モデルと、DTMF信号モデルを備え、前記入力された音信号の照合において、前記話音信号モデルと前記DTMF信号モデルの双方を用いて照合し

言語モデルを備え、前記照合結果と前記言語モデルを用いて音信号の認識を行ない、

前記話音信号区間またはDTMF信号区間のいずれか一方または双方を含む音信号に対する音信号認識を実行することを特徴とする音信号認識方法。

【請求項10】 請求項9に記載の音信号認識方法を含み、前記音信号認識方法を用いた音信号認識結果に基づいて利用者との対話の流れを制御する対話制御方法。

【請求項11】 話音信号区間またはDTMF信号区間の一方または双方を含む入力音信号に対する音信号認識処理を実行する音信号認識プログラムであって

話音信号区間またはDTMF信号区間のいずれか一方または双方を含む音信号を入力する音信号入力処理ステップと、

話音信号モデルと、DTMF信号モデルを備え、前記音信号入力処理ステップにおいて入力された音信号の照合処理において、前記話音信号モデルと前記DTMF信号モデルの双方を用いて照合処理を行なう照合処理ステップと、

単語辞書および文法規則情報を含む言語モデルを備え、前記照合処理ステップにおける照合結果を基に前記言語モデルを用いて音信号の認識を行なう音信号認識処理ステップを備えたことを特徴とする音信号認識プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、利用者が入力した音信号の認識処理を実行する音信号認識システムおよび音信号認識システムを用いた対話制御システムに関する。特に、入力される音信号が、利用者の話音信号のみである場合、タッチトーン式電話システム（プッシュホン式電話システム）から音信号として入力されるDTMF信号（Dual-tone multi frequency signal）による音信号のみの場合、話音信号区間およびDTMF信号区間の双方が混在した音信号である場合のいずれであっても、正しく認識できる音信号認識システムと、当該音信号認識システムによる認識結果を基に利用者との間で対話の流れを制御する対話制御システムに関する。

【0002】

【従来の技術】

コンピュータとのヒューマンインタフェースとして利用者の話音による音声入力が注目されている。従来の音声認識システムは、利用者の話音信号を音声認識し、認識したデータを利用者からの入力データとしてコンピュータに渡す。例えば、パーソナルコンピュータのアプリケーションの口述操作や、テキストデータの口述入力などにおいて用いられ始めている。

【0003】

また、DTMF信号による音信号入力も広く用いられている。このDTMF信号による音信号入力方式は、電話音声案内システムなどに広く用いられている。利用者はタッチトーン式電話システムを用い、タッチトーン電話回線を介してコンピュータと接続し、例えば、利用者はコンピュータから音声データとして電話

回線を介して提供される音声ガイダンスを聞き、当該音声ガイダンスに従ってタッチトーン式電話機の番号ボタンを選択して押下し、コンピュータにデータを入力する。このタッチトーン式電話機の番号ボタン押下により生成されるDTMF信号がDTMF信号である。従来のDTMF信号認識システムは、DTMF信号による音信号を認識し、認識したデータを利用者からの入力データとしてコンピュータに渡す。

#### 【0004】

なお、DTMF信号は、タッチトーン式電話システムにおいて、ボタン押下により生成される信号であり、2つの基本周波数の重畳信号として生成されるものである。図17は、DTMF周波数表の一例を示す図である。この例では、“0”から“9”までの数字と、“A”から“D”までのアルファベットと、“#”マークおよび“\*”マークの合計16のデータが割り当てられている。例えば、数字の“1”に対しては、基本周波数697Hzと基本周波数1209Hzの2つの組み合わせが割り当てられており、タッチトーン式電話機の番号ボタン“1”を押下すると、基本周波数697Hzと基本周波数1209Hzを重畳した合成音信号が生成される。この合成音信号が、数字の“1”に対するDTMF信号となる。

#### 【0005】

一般に、話音信号とDTMF信号の認識処理を比較すると、後者の方が認識率が高く、また、後者の方が処理負荷が小さいなどのメリットがあるが、DTMF信号は表現できるデータ数が少ないので、DTMF信号だけでは対応しきれない複雑なデータ（例えば、利用者の名前など）を入力するため、DTMF信号による入力のみならず利用者の話音による音声入力とを切り換えて、アプリケーションに応じてそれぞれを使い分けるものがある。

#### 【0006】

##### 【発明が解決しようとする課題】

DTMF信号による音信号入力と利用者の話音による音声入力とを併用する従来の電話音声応答システムでは、両入力方式を切り換えることができるものの、DTMF信号によるDTMF信号区間と話音信号区間が混在した音信号を認識処



理することはできない。つまり、従来の電話音声応答システムは、DTMF信号による入力モードと話音による入力モードを切り換えつつ用いるものであり、そのモードの切り換えのため、話音による入力の終了やDTMF信号による入力の終了を示す特定のDTMF信号や、入力モードを他方に切り換える指示を示す特定のDTMF信号（例えば、“#”マークの番号ボタンのDTMF信号）を利用者自らが入力し、モードの切り換え操作を行なう必要がある場合がある。

## 【0007】

図18は、従来のDTMF信号による入力と利用者の話音信号による入力とを併用できる電話音声応答システムの構成例を簡単に示した図である。

## 【0008】

図18において、500は音信号入力部、510は切換部、520は話音信号認識部、530はDTMF信号認識部である。

## 【0009】

音信号入力部500は、外部から入力される音信号を入力する部分である。例えば、電話回線を介して利用者から入力される音信号を受け付ける。

## 【0010】

切換部520は、音信号入力部500から入力された音信号の伝達先を切り換える部分であり、話音信号認識部520またはDTMF信号認識部530のいずれかに渡す。なお、切り換え制御は、例えば、音信号入力部500を介して入力される音信号の中において、入力モードを他方に切り換えるコマンドを示す特定のDTMF信号など特定のDTMF信号を検出した場合に、音信号の伝達先を他方に切り換えるなど方法により行なう。

## 【0011】

話音信号認識部520は、入力された話音信号の音声認識を実行する部分である。

## 【0012】

DTMF信号認識部530は、入力されたDTMF信号の認識を実行する部分である。

## 【0013】

このように、従来の構成では、話音信号認識部 520 と DTMF 信号認識部 530 がそれぞれ独立して設けられ、独立して認識処理を実行している。つまり、DTMF 信号による入力モードでは DTMF 信号認識部 530 を用いて認識処理が実行され、話音による入力モードでは話音信号認識部 520 を用いて認識処理が実行される。

【0014】

なお、従来の構成において、話音信号認識部 520 と DTMF 信号認識部 530 の双方を含み、1 ユニットの認識部とするものもあるが、これは、内部に切換部 510 を含み、認識処理の実行時は、話音信号認識部 520 か DTMF 信号認識部 530 とを切り換えつつ、いずれか一方のみを用いるものであり、本質的に、図 18 と同じ構成である。上記従来の構成によれば、音信号の認識結果として、話音信号のみの認識結果と DTMF 信号のみの認識結果のいずれか一方の結果しか得ることはできない。

【0015】

そのため、従来の電話音声応答システムにおいては以下の問題がある。

【0016】

第 1 には、利用者が話音信号による入力か、DTMF 信号による入力かを切り換える必要があるため、その切り換え操作の負荷が増え、また、どちらのモードで入力していいのか迷う場合もあり、利用者が混乱するという問題がある。

【0017】

第 2 には、電話音声応答システム側が、予定している入力モード以外の入力モードによって音信号の入力を行なうと認識率が低くなり、場合によっては認識不能になるという問題がある。例えば、電話音声応答システム側が、DTMF 信号認識部 530 を用いて音信号認識を行なうことを予定している場合に、利用者が話音で入力すれば、当該話音信号を DTMF 信号認識部 530 では認識できない。

【0018】

第 3 には、話音による音信号区間と DTMF 信号による音信号区間とが混在した音信号を認識することができないため、利用者の利便性に欠けるという問題で

ある。例えば、「登録番号は 1 2 3 4 です」というデータを音信号で入力する場合に、以下のように話音信号区間と D T M F 信号区間が混在した音信号を入力できれば便利である。最初の「登録番号は、」という部分を話音により入力し、続いて「1 2 3 4」という数字の部分をタッチトーン式電話機の電話番号押下によりそれぞれ“1”，“2”，“3”，“4”を示す D T M F 信号により入力し、続いて「です」という部分を話音により入力する。従来の電話音声応答システムは、上記のような話音信号区間と D T M F 信号区間とが混在した音信号の入力が出来ないのも、ユーザ利便性に欠けるものである。

## 【 0 0 1 9 】

第 4 には、電話音声応答システムの設計工数が大きくなり、コスト上昇を招くという問題である。つまり、従来の電話音声応答システムでは、入力モードを正しく誘導するガイダンスが必要となり、対話の流れのアルゴリズムが複雑なものとなり、設計工数の増加に伴うコストの上昇を招いている。

## 【 0 0 2 0 】

そこで、本発明は、入力される音信号が、利用者の話音信号のみの音信号である場合、D T M F 信号のみの音信号である場合、話音信号区間および D T M F 信号区間の双方が混在した音信号である場合のいずれであっても、正しく音信号を認識し、かつ、利用者の入力モードの切り換え操作を不要とする音信号認識システムおよび方法、さらに、音信号認識システムを用いた対話制御システムおよび方法を提供することを目的とする。

## 【 0 0 2 1 】

## 【課題を解決するための手段】

上記目的を達成するため、本発明の音信号認識システムは、話音信号区間又は D T M F 信号区間のいずれか一方または双方を含む音信号を入力する音信号入力部と、話音信号モデルと、D T M F 信号モデルを備え、前記音信号入力部から入力された音信号の照合処理において、前記話音信号モデルと前記 D T M F 信号モデルの双方を用いて照合処理を行なう照合部と、言語モデルを備え、前記照合部の照合結果と前記言語モデルを用いて音信号の認識を行なう音信号認識部を備え、話音信号区間と D T M F 信号区間の一方または双方を含む音信号に対する音信

号認識処理を実行することを特徴とする。

【 0 0 2 2 】

ここで、前記音信号認識部が、認識単位となる一区切りの音信号区間ごとに、前記照合部における前記話音信号モデルを用いた照合結果と前記DTMF信号モデルを用いた照合結果とを比較して照合結果の良い方を選択し、前記音信号区間ごとに選択された音信号認識結果をつないで前記入力音信号の全区間に対する音信号認識結果として統合する統合部を備えるものとしても良い。

【 0 0 2 3 】

上記構成により、本発明の音信号認識システムは、入力される音信号が、利用者の話音信号のみである場合、DTMF信号による音信号のみの場合、話音信号区間およびDTMF信号区間の双方が混在した音信号である場合のいずれであっても、正しく音信号認識することができ、かつ、入力モードの切り換え操作は不要となり、入力モードを正しく誘導するためのシナリオも不要となり、設計工数およびコストの低減を図ることができる。

【 0 0 2 4 】

ここで、言語モデルの単語辞書は、DTMF信号を音信号認識語彙として含むものとすれば、DTMF信号と単語との照合が可能となり、DTMF信号の音信号認識が可能となる。

【 0 0 2 5 】

次に、上記本発明の音信号認識システムにおいて、ガイダンス部を備えるものとしても良い。ガイダンス部により、前記音信号入力部を介して音信号入力を行なう利用者に対して、特定の語彙について、話音による音信号入力とするかDTMF信号による音信号入力とするかをガイダンスすることができる。

【 0 0 2 6 】

ここで、前記統合部が、所定の条件により、特定の語彙について、話音による入力された音信号の誤認識が大きいと検出した場合、ガイダンス部に対して、当該特定語彙について、DTMF信号により音信号を再入力するように求めるガイダンスを表示する指示情報を通知することも可能であり、また、統合部が、話音による音信号に対する照合結果における誤認識率と、DTMF信号による音信号

に対する照合結果における誤認識率を予測、保持し、いずれか一方の誤認識率が所定値よりも高くなった場合、前記ガイダンス部に対して、他方の音信号による入力を促すガイダンスを表示する指示情報を通知することも可能である。ここで、所定の条件とは、音声入力環境、通信環境などにおける S N 比 (signal noise ratio) が所定レベルより悪い場合や、対話の過程で得られた利用者の話音入力の尤度が全般に低い場合に該当することなどが挙げられる。

## 【 0 0 2 7 】

また、上記音信号認識システムを実現する処理プログラムを提供することにより、パーソナルコンピュータなどを用いて、安価かつ手軽に本発明の音信号認識処理を実現することができる。

## 【 0 0 2 8 】

また、上記音信号認識システムを含み、前記音信号認識システムによる音信号認識結果に基づいて利用者との対話の流れを制御する対話制御部を設ければ、音信号認識システムを適用した対話制御システムを提供することができる。

## 【 0 0 2 9 】

## 【発明の実施の形態】

以下、図面を参照しつつ、実施形態 1 から 3 に本発明の音信号認識システムおよび方法を示し、実施形態 4 から 5 に本発明の対話制御システムおよび方法を示し、実施形態 6 に本発明の音信号認識処理ステップおよび対話制御処理ステップを記述したプログラムについて説明する。

## 【 0 0 3 0 】

## (実施形態 1)

本発明の音信号認識システムおよび方法は、DTMF 信号である DTMF 信号の認識処理と話音信号の認識処理の両者を、一つの音信号認識処理において統一的に取り扱うことにより実行するものであり、入力された音信号が、DTMF 信号による音信号のみ、話音による音信号のみ、DTMF 信号区間と話音音声区間が混在した音信号のいずれであっても正しい音信号認識処理が実行できるものである。

## 【 0 0 3 1 】

図 1 は、本発明の実施形態 1 にかかる音信号認識システムの構成および処理の流れの概略を示す図である。

【 0 0 3 2 】

1 0 0 は音信号入力部であり、外部から音信号を入力する部分である。音信号入力部 1 0 0 は、例えば、公衆電話回線に接続され、公衆電話回線から送信される音信号を入力する。また、V o I P 電話システムを利用する場合は、コンピュータネットワークに接続され、ネットワーク上から送信される音信号を入力する。

【 0 0 3 3 】

ここで、入力される音信号としては、利用者の話音信号と D T M F 信号の一方のみの音信号でも良く、また、D T M F 信号区間と話音音声区間が混在した音信号でも良い。

【 0 0 3 4 】

2 0 0 は音信号照合・認識部である。当該音信号照合・認識部 2 0 0 は、入力された音信号が、話音信号または D T M F 信号のどちらか一方と仮定せずに、両者を区別することなく音の信号として統一的に取り扱って照合処理および認識処理を実行するものである。

【 0 0 3 5 】

音信号照合・認識部 2 0 0 の内部の構成は複数通り有り得る。本実施形態 1 における音信号照合・認識部 2 0 0 の内部の構成を図 2 に示す。

【 0 0 3 6 】

図 2 の構成において、音信号照合・認識部 2 0 0 は、音信号分析部 2 1 0、D T M F 信号モデル 2 2 0、話音信号モデル 2 3 0、照合部 2 4 0、言語モデル 2 5 0、認識部 2 6 0 を含んでいる。

【 0 0 3 7 】

音信号分析部 2 1 0 は、音信号入力部 1 0 0 から入力された音信号を、認識単位となる一区切りの音信号区間ごとに分割する処理と、区間ごとに分割した各音信号の特徴量を抽出する処理を行なう部分である。音信号区間の分割処理は、例えば、音信号を一定時間長（フレーム長）に分割する処理とする。特徴量の抽出

処理は、後述するDTMF信号モデルや話音信号モデル作成に採用された特徴量抽出アルゴリズムを用いれば良い。例えば、高速フーリエ変換（FFT）などを用いた特徴量抽出処理を採用し、一定時間長（フレーム長）の音信号に対して一定時間（フレーム周期）ごとに当該処理を実行する。

## 【0038】

DTMF信号モデル220は、各DTMF信号の特徴量を集めたモデル情報である。

## 【0039】

話音信号モデル230は、従来の音声認識同様に認識単位（例：音素、音節、単語）毎に特徴量の分布がどうなっているかをVQやHMM（hidden Markov Model）等を用いて表わすモデル情報である。

## 【0040】

照合部240は、音信号分析部210から渡された各区分ごとの音信号を、DTMF信号モデル220と話音信号モデル230の双方を用いて照合する部分である。本実施形態1は、1つの照合部240によりDTMF信号モデル220と話音信号モデル230の双方を用いて参照するものである。照合処理は、各区分の音信号とモデル内の音素、音節、DTMF音とのマッチングによるスコアを計算し、照合結果を得る。スコアの付け方は自由であるが、例えば、DTMF信号モデル220を用いた照合処理の場合、認識精度が高いため、“1”か“0”かのクリスプとしてスコアを与える。また、話音信号モデル230を用いた照合処理の場合、正規分布を使ったHMMによる音声認識において、ある音素のある状態における出力確率の尤度としてスコアを与える。

## 【0041】

言語モデル250は、単語辞書のみ、単語辞書と文法規則を含むモデル情報である。言語モデル250が保持する単語辞書の例を図3および図4に示す。図3の例の単語辞書は、各々の単語について、単語ID、表記、読み（音の種類で区別）の対応関係を記述したものである。なお、表記を単語IDとすることが可能な場合や、単語IDと表記との対応関係を照合部240が管理している場合などでは、単語辞書中の表記欄は不要である。図4の例の単語辞書は、同じ意味の単

語に対しては統一して同じ単語IDを付し、統一の単語ID、表記、読み（音の種類で区別）の対応関係を記述したものである。言語モデル250が保持する文法規則の例としては、オートマトン文法がある。オートマトン文法の代表的書式としては、BNF法（Backus-Naur Form）がある。

#### 【0042】

認識部260は、各音区間がどういう話音信号またはDTMF信号であるかという尺度として、照合部240からスコアを得て、言語モデル250の単語辞書を参照して、時間方向にDPマッチングなどの探索処理を実行し、入力された全区間のスコアが最も高くなる単語または所定数の上位の単語を探索するものである。この認識結果は、単語辞書が持つ単語IDを用いて表わすことができる。

#### 【0043】

以上の構成により、入力された音信号が、DTMF信号による音信号のみ、話音による音信号のみ、DTMF信号区間と話音音声区間が混在した音信号のいずれであっても、1つの照合部240によりDTMF信号モデル220と話音信号モデル230の双方を用いて参照して照合し、認識部260が照合部240から得たスコアを基に言語モデル250の単語辞書を用いて正しい音信号認識処理を実行する。

#### 【0044】

次に、照合部240において、話音信号区間とDTMF信号区間が混在した音信号が入力された場合に、DTMF信号モデル220と話音信号モデル230の双方を用いた照合処理を詳しく述べる。

#### 【0045】

以下の例は、利用者が音信号認識システムに対して、タッチトーン式電話機のボタン押下により“1”を入力した後、引き続いて話音により“ワシオ”と入力し、対話の中で一括して“1、ワシオ”と入力した場合の音信号認識処理を説明する。

#### 【0046】

図5（a）は、話音信号区間とDTMF信号区間が混在した音信号の概念を示す図であり、図5（b）はDTMF信号スペクトルの例、図5（c）は話音信号



スペクトルの例を示す図である。

【 0 0 4 7 】

図 5 ( a ) に示した音信号は簡単に 2 つの音信号区間 5 1 および 5 2 からなり、5 1 は D T M F 信号である D T M F 信号区間であり、図 5 ( b ) に示すようなスペクトル信号波形を持っている。例えば、利用者がユーザ I D 番号（ここでは“ 1 ”）をタッチトーン式電話機のボタン押下により入力した場合に発生した D T M F 信号音を模式的に示したものである。5 2 は話音信号区間であり、図 5 ( c ) に示すようなスペクトル信号波形を持っている。ここでは、利用者が自らの名前を音声で「ワシオ」と入力した話音信号を模式的に示したものである。

【 0 0 4 8 】

図 5 ( a ) に示した音信号が音信号入力部 1 0 0 から入力され、音信号照合・認識部 2 0 0 に渡される。

【 0 0 4 9 】

また、音信号照合・認識部 2 0 0 の音信号分析部 2 1 0 において、音信号は、音信号区間 5 1 ( 図 5 ( b ) )、音信号区間 5 2 ( 図 5 ( c ) ) に分離される。

【 0 0 5 0 】

( 1 ) 音信号区間 5 1 に対する認識処理

照合部 2 4 0 は、音信号区間 5 1 に対して照合処理を開始する。

【 0 0 5 1 】

照合処理にあたっては、D T M F 信号モデル 2 2 0 を参照した照合処理と、話音信号モデル 2 3 0 を参照した照合処理の双方を並行して参照する。

【 0 0 5 2 】

( a ) D T M F 信号モデル 2 2 0 を参照した照合処理

D T M F 信号モデル 2 2 0 を参照した照合処理の一例は以下のようになる。その流れを図 6 のフローチャートにまとめて示す。

【 0 0 5 3 】

まず、照合部 2 4 0 は、入力された図 5 ( b ) の信号波形のスペクトルからピーク周波数を 2 つ検出する。音信号区間 5 1 の音信号のスペクトル信号波形は、図 5 ( b ) に示したように、2 つのピークを持っているので、この周波数が  $f_1$

、 $f_2$  ( $f_1$ が高い方の周波数、 $f_2$ が低い方の周波数)として検出される(ステップS601)。

【0054】

次に、検出した2つのピーク周波数に対し、図17に示したDTMF周波数表の各周波数成分から、所定の閾値範囲内である周波数成分を探索する(ステップS602)。もし、所定の閾値範囲内である周波数成分が図17のDTMF周波数表の中に見つからない場合(ステップS602:N)、照合部240は、DTMF信号モデル220を参照した照合処理結果としてスコアを“0”とする(ステップS607)。ここでは、音信号区間51では“1”のDTMF信号である例であるので、 $f_1$ が1209Hzで、 $f_2$ が697Hzであることが検知される。

【0055】

ここで、入力音信号の雑音レベルが大きい場合や、入力音信号の波形の歪みが大きい場合、以下のステップS603からステップS605に示す処理をオプションとして行なって、DTMF信号の認識精度を上げることが可能である。

【0056】

第1に、照合部240は、検知した2つのピーク周波数のレベル差が所定の閾値以上であるか否かを調べる(ステップS603)。 $f_1$ のレベル値を $L_1$ 、 $f_2$ のレベル値を $L_2$ とすると、レベル差( $L_2 - L_1$ )が所定の閾値以上であれば(ステップS603:Y)、照合部240は、DTMF信号モデル220を参照した照合処理結果としてスコアを“0”とする(ステップS607)。なぜならば、DTMF信号は、同程度の高いピークを持つ2つの周波数成分が含まれているはずであり、2つのピーク周波数があっても両者の差が所定の閾値よりも大きい場合は、当該音信号が、DTMF信号ではないと推定できるからである。

【0057】

第2に、照合部240は、音信号区間51から3番目に高いピーク( $f_3$ とする)を探索し、 $f_3$ のレベル値 $L_3$ と $f_1$ のレベル値 $L_1$ との差( $L_1 - L_3$ )が所定の閾値以上であるか否かを調べる(ステップS604)。両者のレベル差が所定の閾値以上でない場合であれば(ステップS604:N)、照合部240

は、DTMF信号モデル220を参照した照合処理結果としてスコアを“0”とする（ステップS607）。なぜならば、DTMF信号は、2つの高いピークを持ち、他の周波数成分には高いピークが含まれていないはずであり、 $f_1$ と $f_3$ のピークレベルの差（ $L_1 - L_3$ ）が所定の閾値に満たない場合は、当該音信号が、DTMF信号ではないと推定できるからである。

## 【0058】

第3に、照合部240は、信号波形のスペクトルのうち、 $f_1$ と $f_2$ 付近の周波数レンジ以外の周波数部分、つまり、所定の閾値を $\alpha$ として、 $f_1 \pm \alpha$ および $f_2 \pm \alpha$ の周波数レンジ以外の周波数部分のレベルの平均値（これを $L_4$ とする）を求め、当該平均値 $L_4$ と、 $f_1$ のレベル値 $L_1$ の差（ $L_1 - L_4$ ）が所定の閾値以上であるか否かを調べる（ステップS605）。差（ $L_1 - L_4$ ）が所定の閾値に満たない場合であれば（ステップS605：N）、DTMF信号モデル220を参照した照合処理結果としてスコアを“0”とする（ステップS607）。なぜならば、DTMF信号は、2つの高いピークを持ち、他の周波数成分はすべて、当該2つのピークより十分小さいはずであり、平均値 $L_4$ と、 $f_1$ のレベル値 $L_1$ の差（ $L_1 - L_4$ ）が所定閾値に満たない場合は、当該音信号が、DTMF信号ではないと推定できるからである。

## 【0059】

以上、照合部240は、検知した2つのピーク周波数より、表10のDTMF周波数表に基づいて、音信号区間51の音信号を認識する（ステップS606）。ここでは、音信号区間51が“1”であることが認識され、そのスコア値は大きくなり“1”とされる。

## 【0060】

（b）話音信号モデル230を参照した照合処理

一方、話音信号モデル230を参照した音信号区間51に対する照合処理の一例は以下になる。

## 【0061】

図5（b）のような周波数スペクトルを持つDTMF信号に対して、話音信号モデル230を参照して照合処理を行なうと、話音信号モデル230中には照合

し得る話音候補が見つからない。なぜなら、人の話音信号は、図5(c)に示すのように、広い周波数レンジにまたがる複雑なスペクトルから構成されており、図5(b)のDTMF信号のような2つのピーク周波数を持つ機械音の周波数スペクトルとは大きく異なるため、DTMF信号が話音信号モデルにより照合するとそのスコア値は“0”付近の極めて低い値とされる。

#### 【0062】

照合部240は、上記2つの処理から最も良いスコア値“1”である照合処理の結果を選択し、音信号区間51が、DTMF信号により“1”を表わすものと認識することができる。ここで、照合部240は、音信号区間51がDTMF信号であるか話音信号であるかを区別することなく、正しく認識できたことが分かる。

#### 【0063】

##### (2) 音信号区間52に対する認識処理

次に、照合部240は、音信号区間52に対して照合処理を開始する。

#### 【0064】

音信号区間52に対する照合処理においても、DTMF信号モデル220を参照した照合処理と、話音信号モデル230を参照した照合処理の双方が並行して参照される。

#### 【0065】

##### (a) DTMF信号モデル220を参照した照合処理

DTMF信号モデル220を参照した照合処理の一例は、音信号区間51で用いた図6のフローチャートに従って以下になる。

#### 【0066】

まず、照合部240は、入力された図5(c)の信号波形のスペクトルからピーク周波数を2つ検出する(ステップS601)。音信号区間52の音信号のスペクトル信号波形は図5(c)のようであり、その中から最も大きいレベルを持つもの(例えば、周波数 $f_1'$ でレベルが $L_1'$ (図5(c)には図示せず))と、次に大きいレベルを持つもの(例えば、周波数 $f_2'$ でレベルが $L_2'$ (図5(c)には図示せず))が検出される。

## 【0067】

ここで、音声区間52の信号波形のスペクトルは、図5(c)に示したような広い周波数レンジにまたがる複雑なスペクトルであるので、以下のステップにおいてそのスコア値が“0”となる可能性が極めて高い。つまり、ステップS602における図17に示したDTMF周波数表の周波数成分との照合、ステップS603における $f_1'$ と $f_2'$ のピーク周波数のレベル差チェック、ステップS604における3番目に高いピークのレベル値( $L_3'$ とする)と $f_1'$ のレベル値 $L_1'$ との差( $L_1' - L_3'$ )のチェック、ステップS605における $f_1' \pm \alpha$ および $f_2' \pm \alpha$ の周波数レンジ以外の周波数部分のレベルの平均値(これを $L_4'$ とする)と $f_1'$ のレベル値 $L_1'$ の差( $L_1' - L_4'$ )のチェックにおいて、DTMF信号ではないと推定される可能性が極めて高い。従って、ここでは、S607の処理により、そのスコア値は低くなり“0”とされる。

## 【0068】

(b) 話音信号モデル230を参照した照合処理

一方、話音信号モデル230を参照した音信号区間52に対する照合処理の一例は以下のようなになる。

## 【0069】

図5(c)のような周波数スペクトルを持つ話音信号に対して、話音信号モデル230を参照して照合処理を行なうと、話音信号モデル230の性能が十分であれば、照合し得る話音候補が見つかる。ここでは、音信号区間52は、“ワ”、“シ”、“オ”という3つの話音信号が連続したものとして認識され、そのスコア値は、例えば、正規分布を使ったHMMによる音声認識において当該音素の出力確率の尤度として、明らかに“0”より大きいと判断できる適当な数値とされる。

## 【0070】

以上、照合部240は、照合部240は、上記2つの処理からスコア値が大きい話音信号モデル230を参照した照合処理の結果を選択し、音信号区間52が、“ワシオ”を表わすものと認識する。ここで、照合部240は、音信号区間52がDTMF信号であるか話音信号であるかを区別することなく、正しく認識で

きたことが分かる。

【 0 0 7 1 】

以上、照合部 2 4 0 は、上記 ( 1 ) 音信号区間 5 1 に対する認識処理と、 ( 2 ) 音信号区間 5 2 に対する認識処理とを、何ら装置のモードなどを切り換えることなく、連続して実行できることが分かる。

【 0 0 7 2 】

一方、従来の照合処理によれば、上記 ( 1 ) 音信号区間 5 1 に対する認識処理と、 ( 2 ) 音信号区間 5 2 に対する認識処理において参照するモデルを切り換えなければ正しく照合処理ができない。つまり、DTMF 信号モデル 2 2 0 のみを参照した照合処理を実行した場合、音信号区間 5 1 は正しく “ 1 ” と認識できても音信号区間 5 2 は “ ワシオ ” と正しく認識できない。つまり、利用者が “ 1、ワシオ ” と DTMF 信号区間 5 1 と話音信号区間 5 2 を混在して入力した場合、正しく認識できないこととなる。同様に、話音信号モデル 2 3 0 のみを参照した照合処理を実行した場合、音信号区間 5 1 は正しく “ 1 ” と認識できず、音信号区間 5 2 のみ “ ワシオ ” と正しく認識するのみである。

【 0 0 7 3 】

以上、本実施形態 1 の音信号認識システムによれば、 1 つの照合処理の中で、DTMF 信号モデルと話音信号モデルを参照することにより、DTMF 信号認識処理と話音信号認識処理の両者を一つの音信号認識処理において統一的に取り扱い、入力された音信号が、DTMF 信号による音信号のみ、話音による音信号のみ、DTMF 信号区間と話音音声区間が混在した音信号のいずれであっても正しい音信号認識処理が実行できる。

【 0 0 7 4 】

( 実施形態 2 )

本発明の実施形態 2 の音信号認識システムおよび方法は、DTMF 信号モデルを参照した DTMF 信号照合処理と、話音信号モデルを参照した話音信号照合処理を並行して実行し、両者の結果を統合することにより、一つの音信号認識処理として統一的に取り扱うものであり、入力された音信号が、DTMF 信号による音信号のみ、話音による音信号のみ、DTMF 信号区間と話音音声区間が混在し

た音信号のいずれであっても正しい音信号認識処理が実行できるものである。

【0075】

本発明の実施形態2にかかる音信号認識システムの構成は、実施形態1で説明した図1と同様、音信号入力部と音信号照合・認識部を備えるものであるが、音信号照合・認識部200aが実施形態1で説明した音信号照合・認識部200とは異なる構成を備えている。

【0076】

本実施形態2における音信号照合・認識部200aの内部の構成を図7に示す。

【0077】

図7の構成において、音信号照合・認識部200aは、音信号分析部210、DTMF信号モデル220、話音信号モデル230、DTMF信号照合部240a、話音信号照合部240b、統合部270、言語モデル250、認識部260を含んでいる。

【0078】

音信号分析部210、DTMF信号モデル220、話音信号モデル230、言語モデル250、認識部260の各要素は実施形態1と同様であるのでここでの説明は省略する。

【0079】

本実施形態2の照合部は、DTMF信号モデル220を参照して照合を行うDTMF信号照合部240aと話音信号モデル230を参照して照合を行う話音信号照合部240bの2つを用いて照合を行うものである。

【0080】

DTMF信号照合部240aは、音信号分析部210から渡された各区分ごとの音信号を、DTMF信号モデル220を用いて照合する部分である。照合処理は、各区分の音信号とモデル内の音素、音節、DTMF音とのマッチングによるスコアを計算し、照合結果を得るものである。スコアの付け方は自由であるが、例えば、DTMF信号モデル220を用いた照合処理の場合、認識精度が高いため、“1”か“0”かのクリスプとしてスコアを与える。

## 【 0 0 8 1 】

話音信号照合部 2 4 0 b は、音信号分析部 2 1 0 から渡された各区分ごとの音信号を、話音信号モデル 2 3 0 を用いて照合する部分である。照合処理は、各区分の音信号とモデル内の音素、音節とのマッチングによるスコアを計算し、照合結果を得るものである。スコアの付け方は自由であるが、音声の発声は、DTMF 信号の場合に比べてバリエーションに富むため、例えば、正規分布を使った HMM による音声認識処理とし、話音信号照合部 2 4 0 b は、ある音素のある状態の出力確率として尤度を出力するものとする。正規分布であるので、図 8 のように分散が大きい場合、出力確率は最大値であっても、1 よりもかなり小さい数値となる。そこで、ダイナミックレンジの確保のために対数尤度とする。また、整数倍して対数尤度を整数化すると、後続の演算処理の高速化を図ることができる。

## 【 0 0 8 2 】

統合部 2 7 0 は、DTMF 信号照合部 2 4 0 a による照合結果と話音信号照合部 2 4 0 b による照合結果を統合する部分である。統合部 2 7 0 を備える理由は以下のとおりである。

## 【 0 0 8 3 】

DTMF 信号照合部 2 4 0 a による照合結果の数値レンジと話音信号照合部 2 4 0 b による処理結果の数値レンジは図 9 に示すように全く異なることが考えられる。この場合に、DTMF 信号照合部 2 4 0 a による照合結果と話音信号照合部 2 4 0 b による照合結果を単に比較して結果の良い方を採用する方式とすると、例えば、話音信号照合部 2 4 0 b による照合結果において照合確率が高いとして良いスコア値を出し、DTMF 信号照合部 2 4 0 a による照合結果においては照合確率が低く、悪いスコア値を出している場合であっても、両者の数値レンジが異なるため、前者の結果より後者の結果の方が照合確率が高いとして後者の結果が採用され、誤認識が起こる可能性がある。そこで、統合部 2 7 0 により両者のレンジ差を調整するのである。統合部 2 7 0 によるレンジ調整を経て両者の照合結果を比較し、スコア値の高い方を選択することにより、正しい認識結果を得ることができる。



## 【 0 0 8 4 】

なお、統合部 2 7 0 の出力に基づいて、認識部 2 6 0 が照合部 2 4 0 から得たスコアを基に言語モデル 2 5 0 の単語辞書を用いて正しい音信号認識処理を実行する処理は実施形態 1 と同様である。

## 【 0 0 8 5 】

以上、本実施形態 2 の音信号認識システムによれば、DTMF 信号モデルを参照した DTMF 信号照合処理と、話音信号モデルを参照した話音信号照合処理を並行して実行し、両者結果を統合することにより、一つの音信号認識処理として統一的に取り扱うものであり、入力された音信号が、DTMF 信号による音信号のみ、話音による音信号のみ、DTMF 信号区間と話音音声区間が混在した音信号のいずれであっても正しい音信号認識処理が実行できる。

## 【 0 0 8 6 】

## (実施形態 3)

本発明の実施形態 3 の音信号認識システムおよび方法は、実施形態 1 の構成の照合部に対して外部から参照するモデルの選択を指示する仕組みを追加したものである。

## 【 0 0 8 7 】

本発明の実施形態 3 にかかる音信号認識システムの構成は、実施形態 1 で説明した図 1 と同様であり、音信号入力部と音信号照合・認識部を備えるものであるが、音信号照合・認識部 2 0 0 b が実施形態 1 で説明した音信号照合・認識部 2 0 0 とは異なる構成を備えている。

## 【 0 0 8 8 】

本実施形態 3 における音信号照合・認識部 2 0 0 b の内部の構成を図 1 0 に示す。図 1 0 の構成において、音信号照合・認識部 2 0 0 b は、音信号分析部 2 1 0、DTMF 信号モデル 2 2 0、話音信号モデル 2 3 0、照合部 2 4 0 c、言語モデル 2 5 0、認識部 2 6 0 を含んでいる。照合部 2 4 0 c は外部からモデル選択信号の入力を受け付ける入力部分を備えている。

## 【 0 0 8 9 】

照合部 2 4 0 c は、モデル選択信号の入力を受け付け、照合処理において用い

るモデルを選択する。この例では、DTMF信号モデル220、話音信号モデル230のいずれか一方のみまたは双方の選択が可能となっている。

【0090】

例えば、音信号入力環境や通信環境の影響などにより、話音信号による入力に対して誤認識が多い場合は、話音による入力を停止し、DTMF信号による入力のみに切り換えた方が好ましいと言える。例えば、利用者が入力した話音が正しく認識されないことが多いと感じた場合や、アプリケーション側が想定しうる利用者の応答内容とは異なる内容の入力が多いと判断した場合に、利用者に対して話音入力を停止し、DTMF信号による入力を誘導するとともに、照合部240cに対するモデル選択信号を与え、DTMF信号モデル220のみを選択した構成とする。この構成によれば、音信号認識システムの照合部240cは、DTMF信号モデルのみを参照し、話音信号モデル230を参照しないものとなる。

【0091】

また逆に、DTMF信号による入力に対して誤認識が多い場合は、DTMF信号による入力を停止し、話音による入力のみに切り換えた方が好ましいと言える。この場合も同様に、利用者に対してDTMF信号による入力を停止し、話音による入力を誘導するとともに、照合部240cに対するモデル選択信号を与え、話音信号モデル230のみを選択した構成とすれば良い。

【0092】

以上、本実施形態3の音信号認識システムは、実施形態1の構成の照合部に対して外部から参照するモデルの選択を指示する仕組みを追加することにより、音信号入力環境や通信環境の影響を考慮し、DTMF信号モデル220、話音信号モデル230のいずれか一方のみまたは双方を選択することができる。

【0093】

(実施形態4)

本発明の実施形態4は、実施形態1から実施形態3に示した音信号認識システムを適用した対話制御システムである。特に、自動電話対応システムにより利用者からの商品発注を受け付けるアプリケーションに適用した対話制御システムを説明する。

【 0 0 9 4 】

図 1 1 は、実施形態 4 にかかる本発明の音信号認識システムを適用した対話制御システムの構成例を示す図である。

【 0 0 9 5 】

図 1 1 において、音信号入力部 1 0 0 および音信号照合・認識部 2 0 0 は、実施形態 1 から実施形態 3 に示したものと同様である。なお、音信号照合・認識部 2 0 0 は、実施形態 2 で説明した音信号照合・認識部 2 0 0 a，実施形態 3 で説明した音信号照合・認識部 2 0 0 b であっても良い。

【 0 0 9 6 】

なお、この例では、音信号照合・認識部 2 0 0 の言語モデル 2 5 0 の使用する単語辞書は、図 4 に示すタイプのものであるとする。

【 0 0 9 7 】

本発明の実施形態 4 の対話制御システムは、さらに、対話管理部 3 0 0、利用者 ID 情報管理部 3 1 0、商品 ID 情報管理部 3 2 0、シナリオ管理部 3 3 0、応答音声出力部 3 4 0 を備えている。なお、この例では、アプリケーションシステムは商品発注システム 4 0 0 となる。

【 0 0 9 8 】

利用者 ID 情報管理部 3 1 0 はユーザ ID と氏名の対応情報を管理する部分である。

【 0 0 9 9 】

商品 ID 情報管理部 3 2 0 は、商品 ID と商品名の対応情報を管理する部分である。

【 0 1 0 0 】

シナリオ管理部 3 3 0 は、対話をどう進めるかというシナリオを管理している。対話の各段階において想定される利用者からの入力情報と当該入力情報に対する応答となる出力情報、対話の各段階において利用者に対して入力を求めるべき項目情報と当該項目情報の入力を求める質問となる出力情報などもシナリオに含まれる。

【 0 1 0 1 】

応答音声出力部 3 4 0 は対話管理部の指定に応じた内容を合成音声／蓄積音声でユーザに返す部分である。

#### 【 0 1 0 2 】

対話管理部 3 0 0 は、利用者への応対、利用者との対話の流れを制御する部分である。対話管理部 3 0 0 はシナリオ管理部 3 3 0 の持つシナリオに従って利用者との間の対話を進め、利用者から発注を受けて、商品発注システム 4 0 0 に発注内容を送信する。

#### 【 0 1 0 3 】

この例では利用者との対話は音信号で行なう。対話制御システム側から利用者に対する対話は、対話内容を表わす指令信号を応答音声出力部 3 4 0 に送り、応答音声出力部 3 4 0 が音声信号に変換して利用者側の機器が備えるスピーカから音声として出力する。利用者に対し、商品に関する情報の提供などを音声で行なったり、利用者に対し、利用者 I D 情報や商品発注情報などの音信号入力に関するガイダンスを行なったりする。

#### 【 0 1 0 4 】

一方、利用者から対話制御システム側に対する対話は、利用者側に備えるタッチトーン電話機器に対して、話音入力や D T M F 信号入力により行なう。

#### 【 0 1 0 5 】

図 1 2 は、利用者と本実施形態 4 の対話制御システムとの間で交わされる対話の流れの一例を示す図である。図 1 2 の例のように、利用者は、D T M F 信号区間のみからなる入力（例えば、図 1 2 の U 1 の入力）、話音信号区間のみからなる入力（例えば、図 1 2 の U 2 の入力）、D T M F 信号区間と話音信号区間が混在した入力（例えば、図 1 2 の U 3 の入力）を行なうことができる。対話制御システムは、本発明の音信号認識システムの働きにより、それら 3 者のいずれの音信号入力であっても正しく認識することができる。

#### 【 0 1 0 6 】

図 1 3 は、商品発注に関するアプリケーションにおいて、利用者と本実施形態 4 の対話制御システムとの間で交わされる対話の流れの一例を示すフローチャートである。図 1 4 は、図 1 3 のフローチャートのうち、ステップ S 1 3 0 1 の利

用者 I D 情報の取得のステップを詳しく示したフローチャートである。

【0107】

図 1 4 の例に示すように、入力された利用者 I D 情報が、利用者 I D 情報管理部 3 1 0 の管理する I D 情報の中で該当するものがあるまで入力が繰り返され、その入力は話音信号でも D T M F 信号でも認識可能であり、その認識結果は単語 I D で対話管理部 3 0 0 に返されるので、以下の利点が得られる。

【0108】

第 1 には、話音信号でも D T M F 信号でも認識可能であるので、従来システムのように音声の入力待ちをするのか、D T M F 信号の入力処理待ちをするのかといった分岐選択処理ステップが不要となる。

【0109】

第 2 には、利用者からの入力を話音信号、D T M F 信号のどちらかに限定させるガイダンス処理ステップ、入力待ち処理ステップが不要となる。

【0110】

第 3 には、利用者からの入力が話音信号、D T M F 信号のどちらが選択されたかによって、対話システムの音信号認識システムによる認識処理を分岐させる必要がない。

【0111】

なお、図 1 4 は、図 1 3 のフローチャートのうち、ステップ S 1 3 0 1 の利用者 I D 情報の取得のステップを詳しく示すものであるが、同様にして、名前情報取得処理（ステップ S 1 3 0 2）、注文情報取得処理（ステップ S 1 3 0 3）、住所情報取得処理（ステップ S 1 3 0 4）に関しても、図 1 4 のフローチャートと同様の流れとすることができ、上記効果を得ることができる。

【0112】

（実施形態 5）

実施形態 5 にかかる本発明の音信号認識システムを適用した対話システムは、音声入力環境、通信環境などにおける S N 比（signal noise ratio）が所定レベルより悪い場合や、対話の過程で得られた利用者の話音入力の尤度が全般に低い場合など、その場の状況に応じて、話音入力に代え、D T M F 信号での入力を促

す対話システムである。なお、この例でも、対話制御システムを、自動電話応対システムにより利用者からの商品発注を受け付けるアプリケーションに適用した例を説明する。

#### 【0113】

SN比が所定レベルより悪い場合に、利用者に対してDTMF信号入力を誘導する場合のシステム構成例を図15に示す。

#### 【0114】

図15において、音信号入力部100および音信号照合・認識部200bは、実施形態3に示したものと同様である。なお、この例では、音信号照合・認識部200bの言語モデル250の使用する単語辞書は、図4に示すタイプのものであるとする。また、対話管理部300、利用者ID情報管理部310、商品ID情報管理部320、シナリオ管理部330、応答音声出力部340、商品発注システム400は実施形態4で説明したものと同様である。

#### 【0115】

実施形態5の対話制御システムは、さらに、SN比算出部350を備えている。SN比算出部350は、音信号入力部100から入力された音信号を受け、そのSN比を計算し、対話管理部に出力する部分である。なお、SN比算出部350が音信号照合・認識部200b内部に包含された構成も可能である。

#### 【0116】

対話管理部300は、SN比算出部350から受け取ったSN比の値がある閾値以上であればSN比が悪いと判断する。対話管理部300は、SN比が悪いと判断した状況下で、対話のシナリオにおいて、利用者に何らかの入力を促すフェーズに至った場合、利用者に対してDTMF信号入力を誘導する。例えば、応答音声出力部340を通じて、DTMF信号入力を促すメッセージとして「雑音が大きいですので、音声入力よりプッシュボタン入力がお勧めです。」と出力する。

#### 【0117】

また、対話の過程で得られた利用者の話音入力の尤度が全般に低い場合も同様に誘導することができる。

## 【0118】

以上、本実施形態5の対話制御システムによれば、利用者から入力された話音信号のSN比が所定レベルより悪い場合や、対話の過程で得られた利用者の話音入力の尤度が全般に低い場合など、その場の状況に応じてDTMF信号での入力を促すことができ、誤認識を少なくし、対話の流れをスムーズにすることができる。

## 【0119】

## (実施形態6)

本発明の音信号認識システムおよび方法、さらに、本発明の音信号認識システムを適用した対話制御システムおよび方法は、上記に説明した構成を実現する処理ステップを記述したプログラムとして記述することができ、当該プログラムをコンピュータに読み取らせることにより、本発明の音信号認識処理を実行することができる。本発明の音信号認識システムを実現する処理ステップを備えたプログラムは、図16に図示した例のように、CD-ROM1002やフレキシブルディスク1003等の可搬型記録媒体1001だけでなく、ネットワーク上にある記録装置内の記録媒体1000や、コンピュータのハードディスクやRAM等の記録媒体1005に格納して提供することができ、ネットワークからダウンロードすることもできる。プログラム実行時には、プログラムはコンピュータ1004上にローディングされ、主メモリ上で実行される。

## 【0120】

なお、本発明の音信号認識システムは、電話回線のみならずVoIPを使ったIP電話のように電話回線を模擬したネットワーク通信システムや、DTMF音発信機能と音声入力（マイク入力）機能を有するリモコン装置などにおいても適用することができる。

## 【0121】

## 【発明の効果】

本発明の音信号認識システムおよび対話制御システムによれば、話音信号区間とDTMF信号区間が混在した音信号を入力することができ、利用者は、話音入力とDTMF信号入力とを区別することなく、自由に入力することができる。

【0122】

また、本発明の音信号認識システムおよび対話制御システムによれば、対話処理時間の短縮、認識率の向上など利用者の使い勝手向上が見込まれる。さらに、対話制御の簡素化、対話処理に関する設計工数削減、それに伴うコスト削減といった効果も期待できる。

【図面の簡単な説明】

【図1】 本発明の実施形態1の音信号認識システムの構成および処理の流れの概略を示す図

【図2】 実施形態1の音信号照合・認識部200の内部の構成を示す図

【図3】 言語モデル250が保持する単語辞書の例を示す図

【図4】 言語モデル250が保持する他の単語辞書の例を示す図

【図5】 (a)は、話音信号区間とDTMF信号区間が混在した音信号の概念を示す図、(b)はDTMF信号スペクトルの例を示す図、(c)は話音信号スペクトルの例を示す図

【図6】 DTMF信号モデル220を参照した照合処理の流れを示すフローチャート

【図7】 実施形態2の音信号照合・認識部200aの内部の構成を示す図

【図8】 正規分布に従った分散が大きい場合の出力確率尤度を示す図

【図9】 DTMF信号照合部240aによる照合結果の数値レンジと話音信号照合部240bによる処理結果の数値レンジが異なる様子を説明する図

【図10】 実施形態3の音信号照合・認識部200bの内部の構成を示す図

【図11】 実施形態4にかかる、本発明の音信号認識システムを適用した対話制御システムの構成例を示す図

【図12】 利用者と実施形態4の対話制御システムとの間で交わされる対話の流れの一例を示す図

【図13】 商品発注に関するアプリケーションにおいて、利用者と本実施形態4の対話制御システムとの間で交わされる対話の流れの一例を示すフローチャート

【図14】 図13のフローチャートのうち、ステップS1301の利用者I



D情報の取得のステップを詳しく示したフローチャート

【図15】 実施形態5にかかる、SN比が所定レベルより悪い場合に、利用者に対してDTMF信号入力を誘導する対話制御システムの構成例を示す図

【図16】 本発明の実施形態4の音信号認識システムを実現する処理プログラムを記録した記録媒体の例を示す図

【図17】 従来のDTMF周波数表の一例を示す図

【図18】 従来のDTMF信号による入力と利用者の話音による入力とを併用できる電話音声応答システムの構成例を簡単に示した図

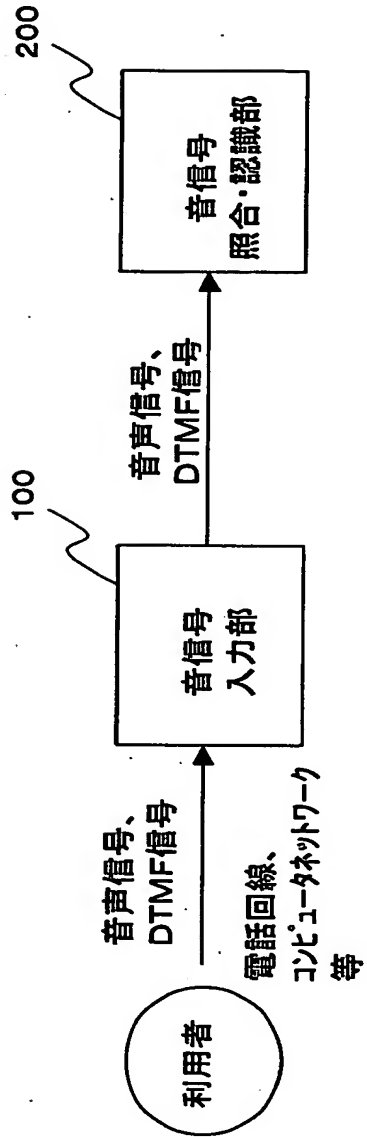
【符号の説明】

- 100 音信号入力部
- 200, 200a, 200b 音信号照合・認識部
- 210 音信号分析部
- 220 DTMF信号モデル
- 230 話音信号モデル
- 240 照合部
- 240a DTMF信号照合部
- 240b 話音信号照合部
- 250 言語モデル
- 260 認識部
- 270 統合部
- 300 対話管理部
- 310 利用者ID情報管理部
- 320 商品ID情報管理部
- 330 シナリオ管理部
- 340 応答音声出力部
- 350 SN比算出部
- 400 商品発注システム
- 1000 回線先のハードディスク等の記録媒体
- 1001 CD-ROMやフレキシブルディスク等の可搬型記録媒体

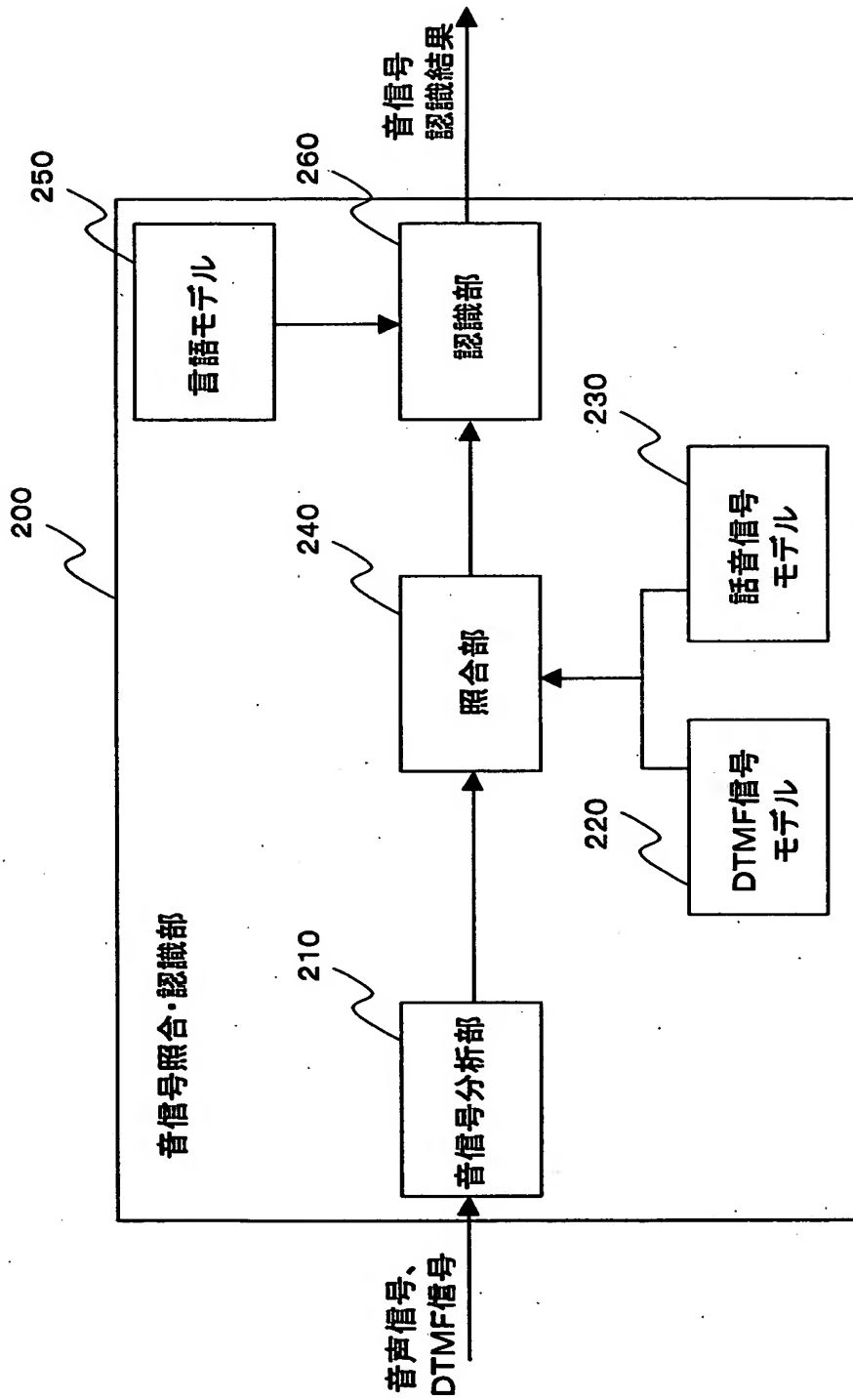
- 1 0 0 2   CD-ROM
- 1 0 0 3   フレキシブルディスク
- 1 0 0 4   コンピュータ
- 1 0 0 5   コンピュータ上のRAM／ハードディスク等の記録媒体

【書類名】 図面

【図 1】



【図 2】



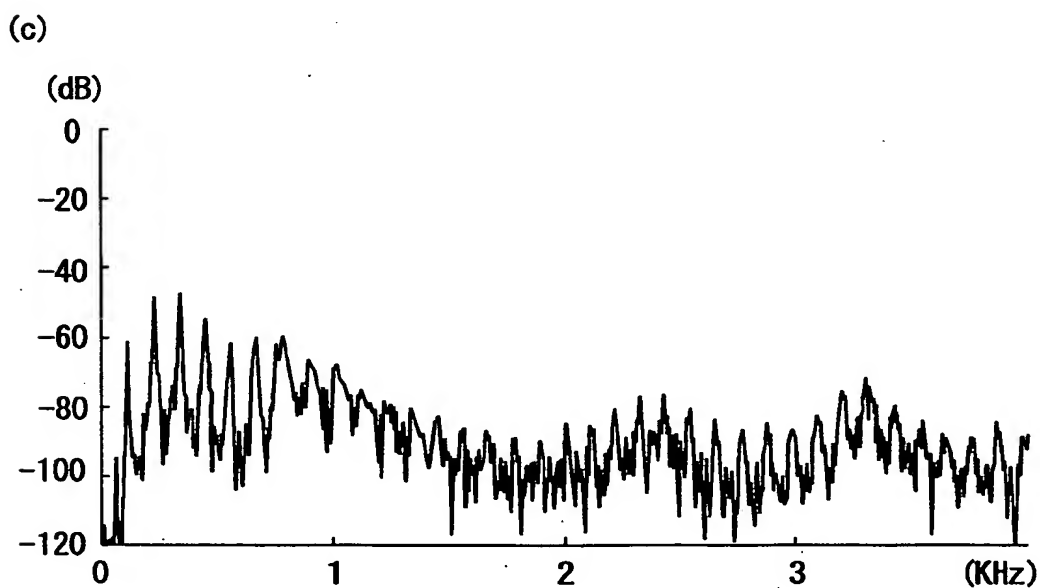
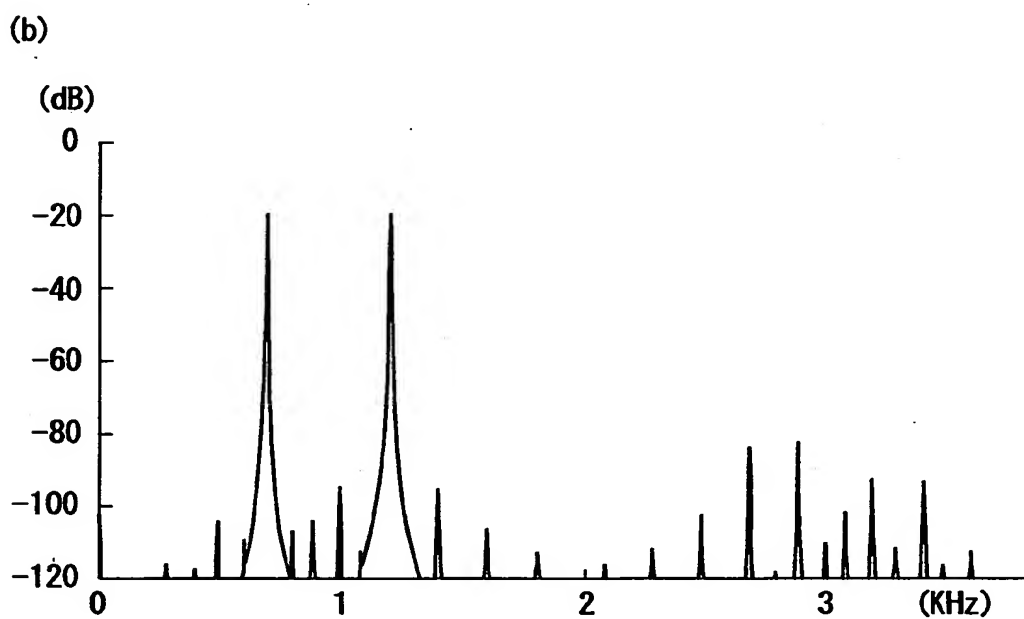
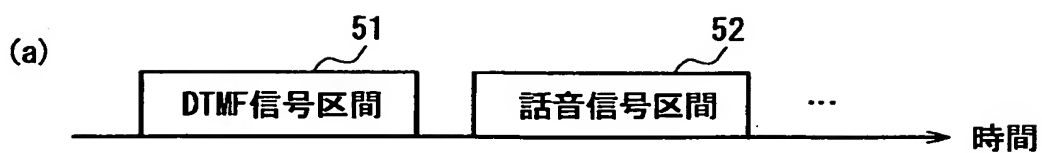
【図 3】

単語ID	表記	読み(音声)
1	0	ゼロ
2	0	しい
3	0	DTMF-0
4	1	イチ
5	1	DTMF-1
6	2	ニ
7	2	DTMF-2
...	...	...
35	はい	ハイ
36	イエス	イエス
37	はい	DTMF-*
38	いいえ	イイエ
39	ノー	ノー
40	いいえ	DTMF-#
...	...	...

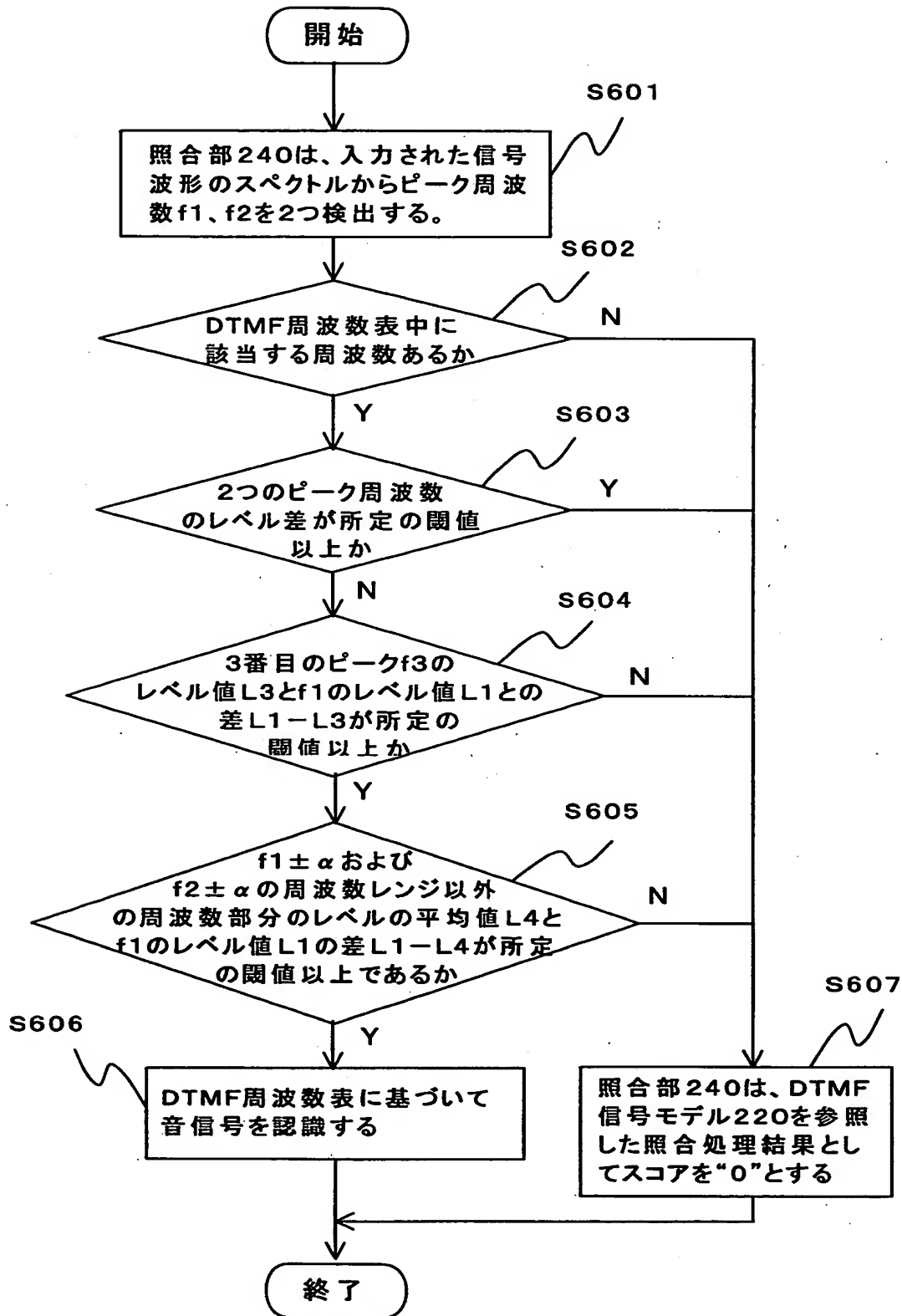
【図 4】

単語ID	表記	読み(音声)
1	0	ゼロ
1	0	レイ
1	0	DTMF-0
2	1	イチ
2	1	DTMF-1
3	2	ニ-
3	2	DTMF-2
...	...	...
4	はい	ハイ
4	イエス	イエス
4	はい	DTMF-*
5	いいえ	イイエ
5	ノー	ノー
5	いいえ	DTMF-#
...	...	...

【図 5】

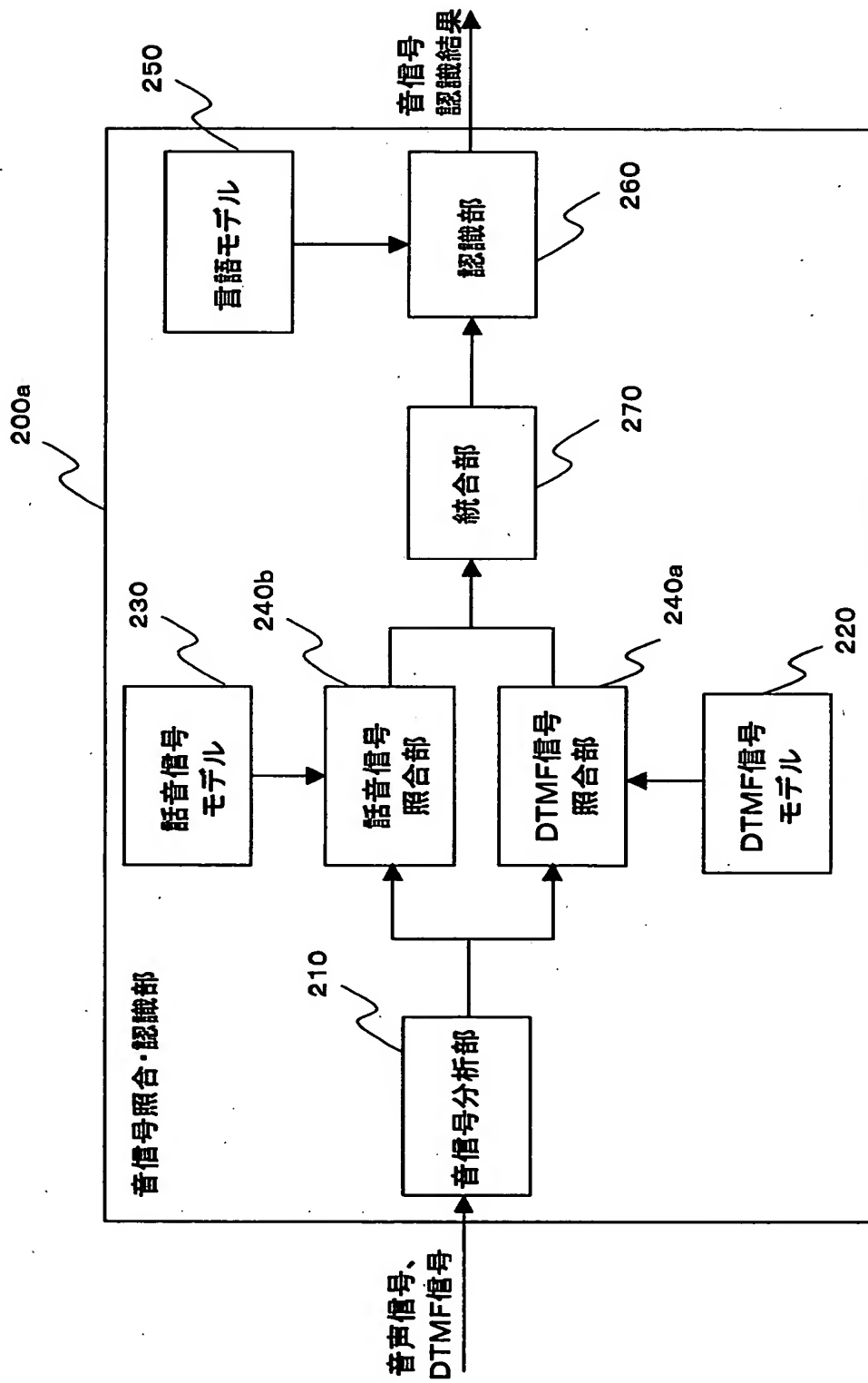


【図6】

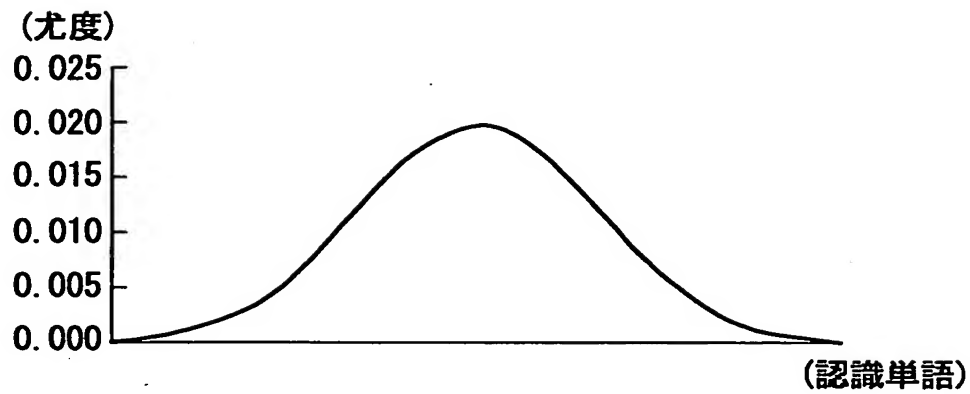




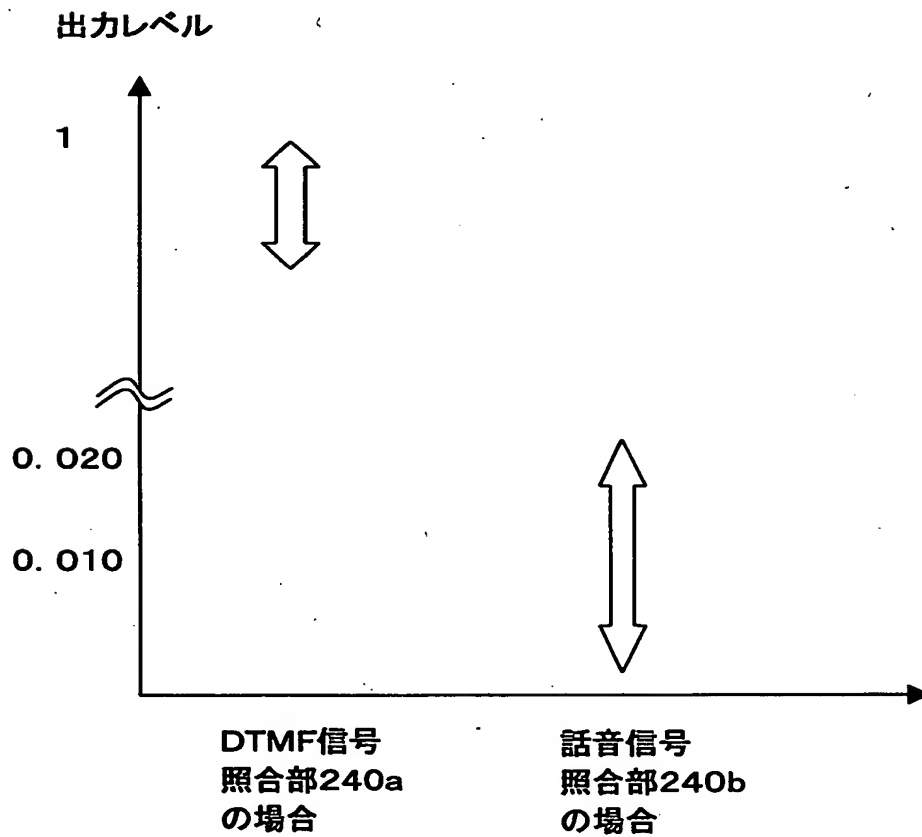
【図 7】



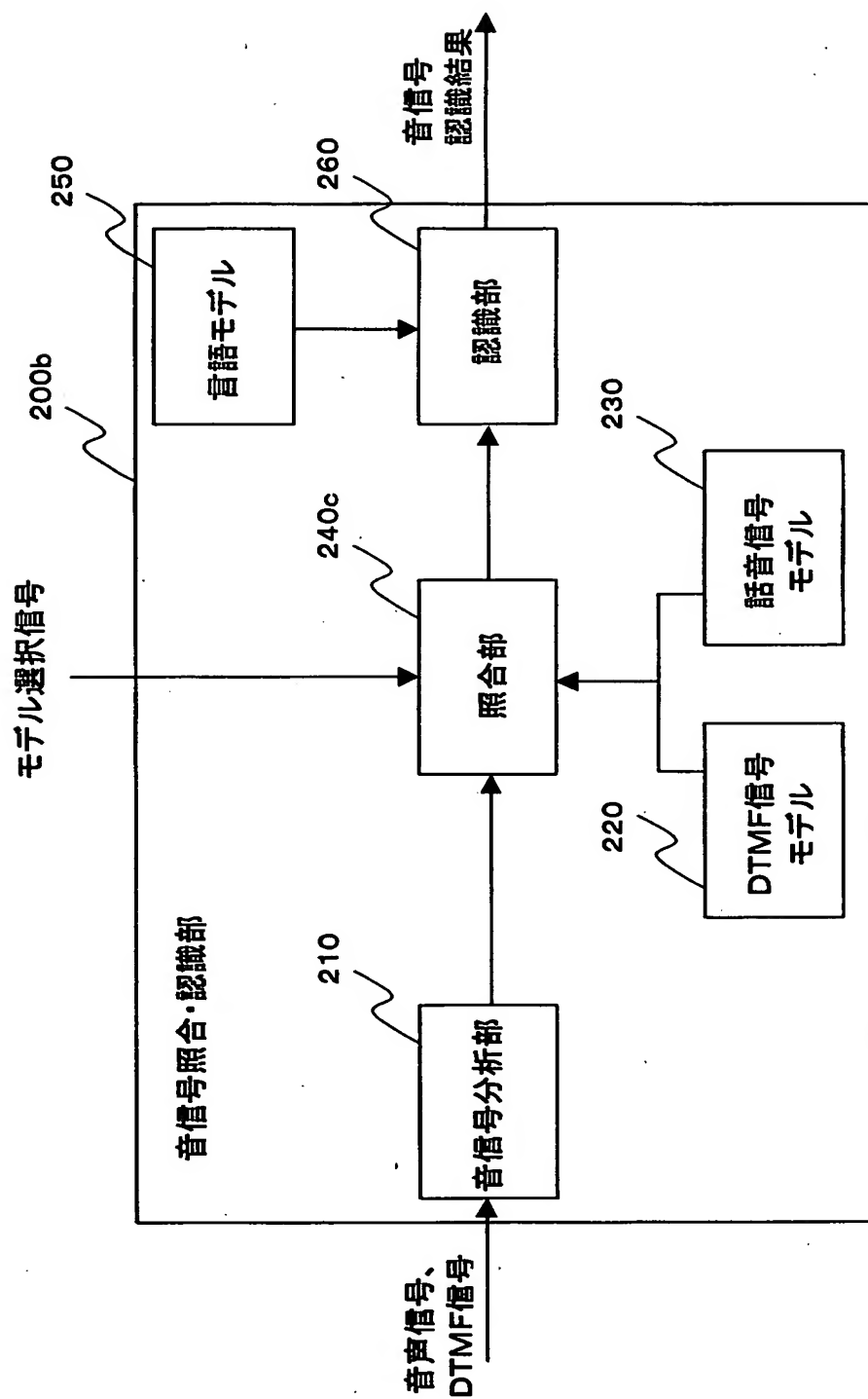
【図 8】



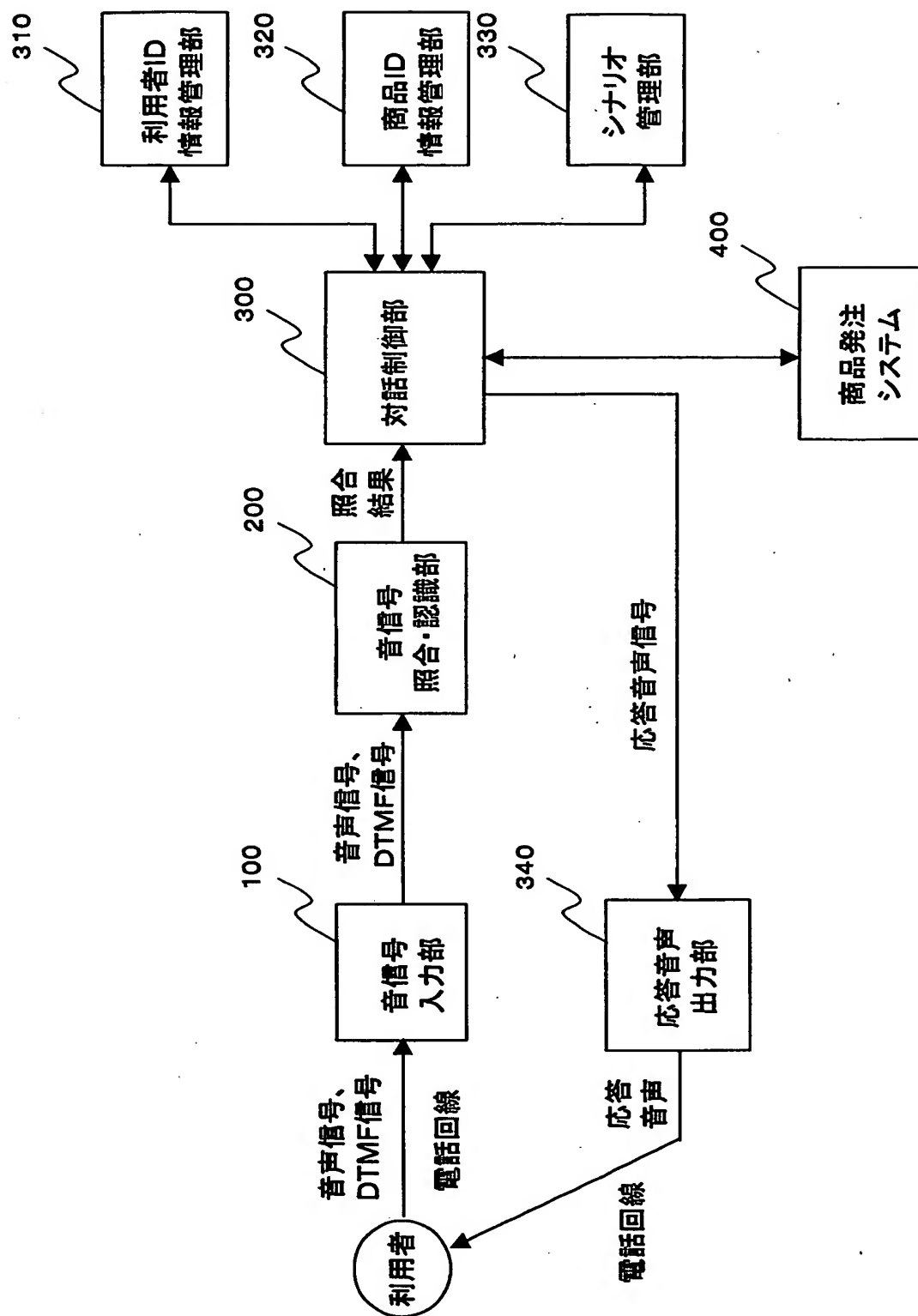
【図 9】



【図10】



【図11】

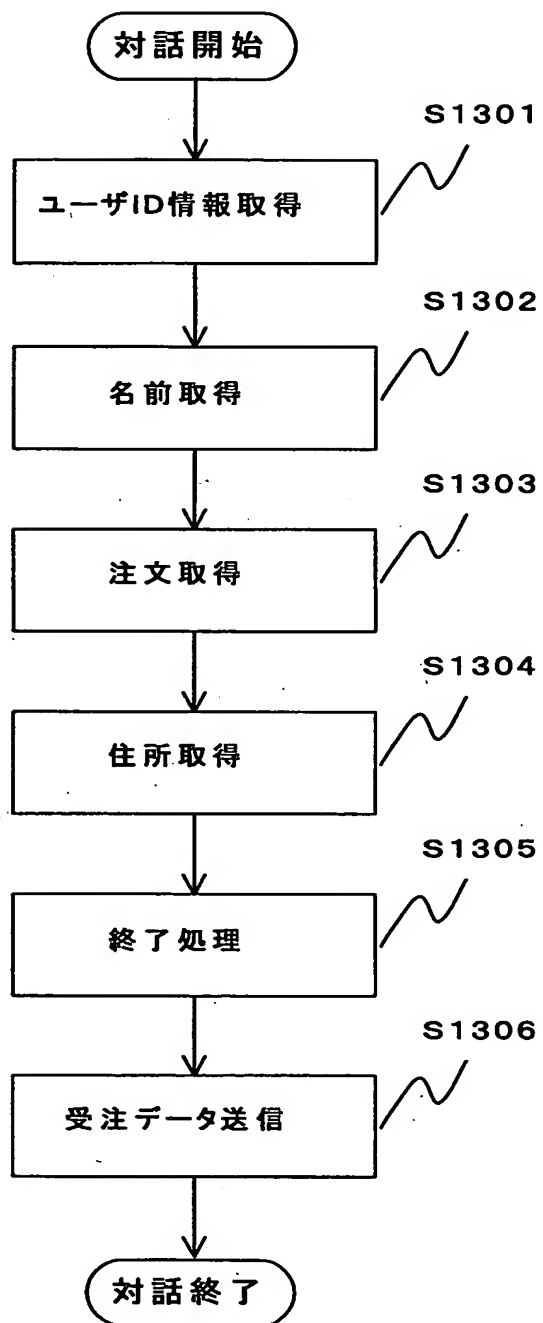


【図 1 2】

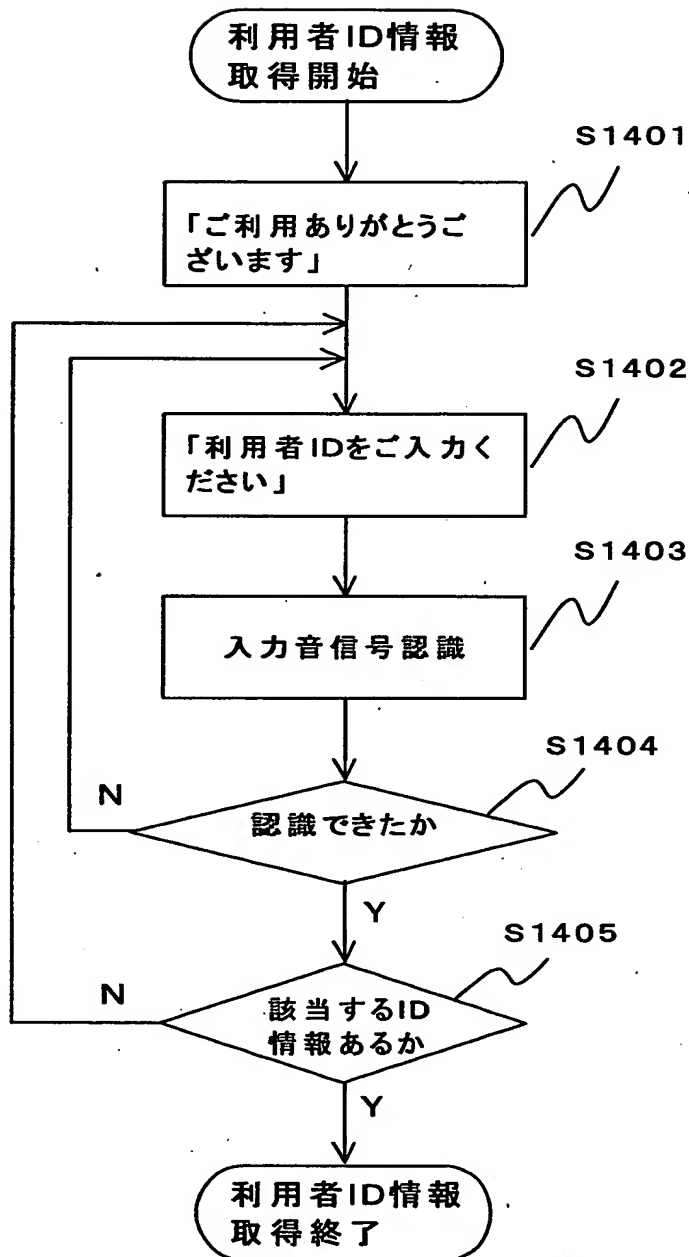
対話例: (U: ユーザの入力、S: 対話システムのからの応答)

- S1: ご利用ありがとうございます。
- S2: ユーザIDをどうぞ。
- U1: ピポピポ(1212を入力)
- S3: お名前をどうぞ
- U2: 富士通太郎
- S4: ユーザ ID 1212 の富士通太郎様ですね。ご希望商品をどうぞ。
- U3: 商品番号ピポパポ(3821 を入力)を1つお願いします
- S5: 商品番号3821のカーフックスを1つですね。
- U4: ピ(DTMF で\*([はい]と対応付けされている))
- S6: ご住所をどうぞ。
- U5: 川崎市中原区上小田中4丁目ポ(1を入力)のポ(1を入力)
- S6: 川崎市中原区上小田中4丁目1の1ですね。ご注文を承りました。
- S7: ご利用ありがとうございました。

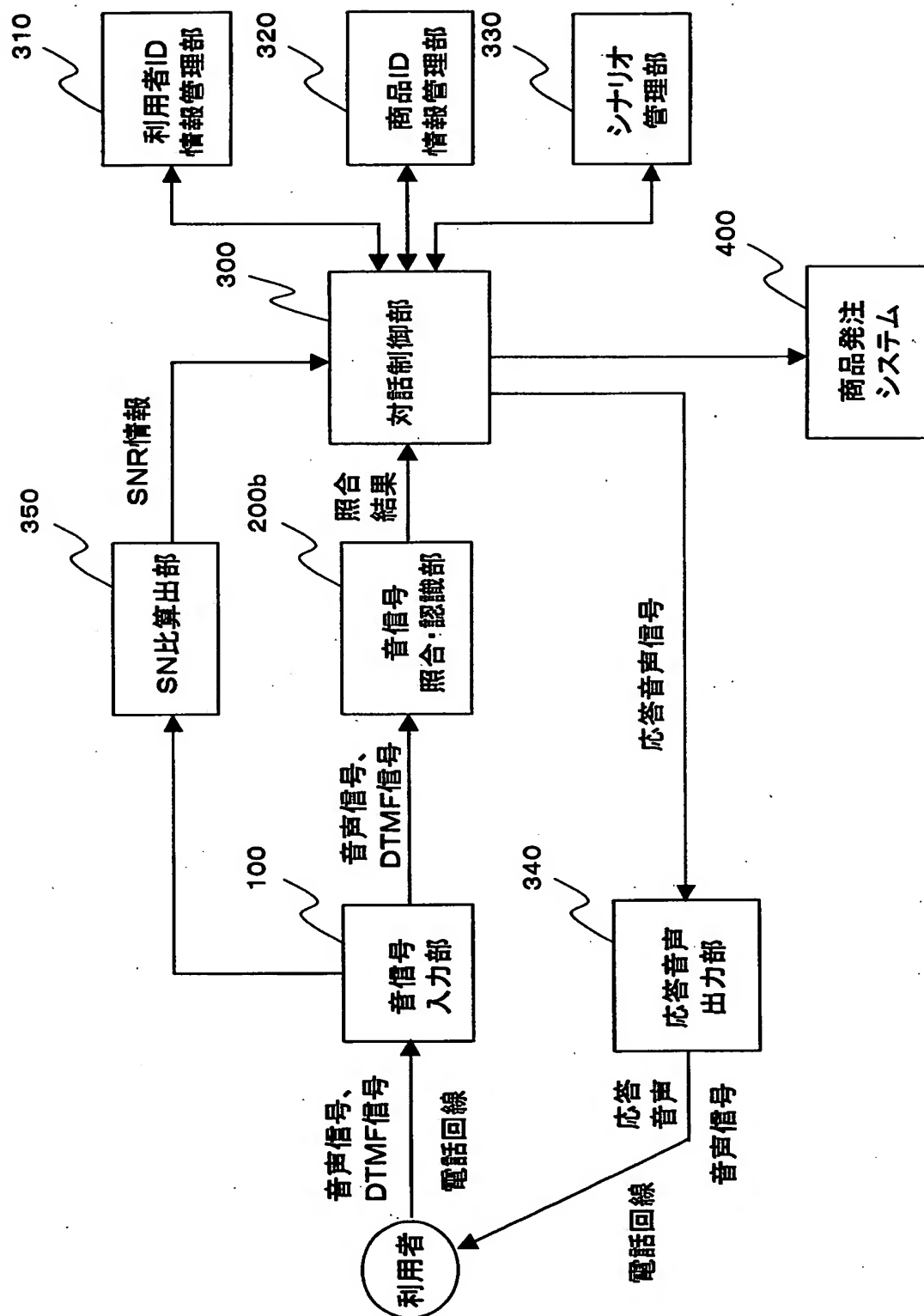
【図 1 3】



【図 14】

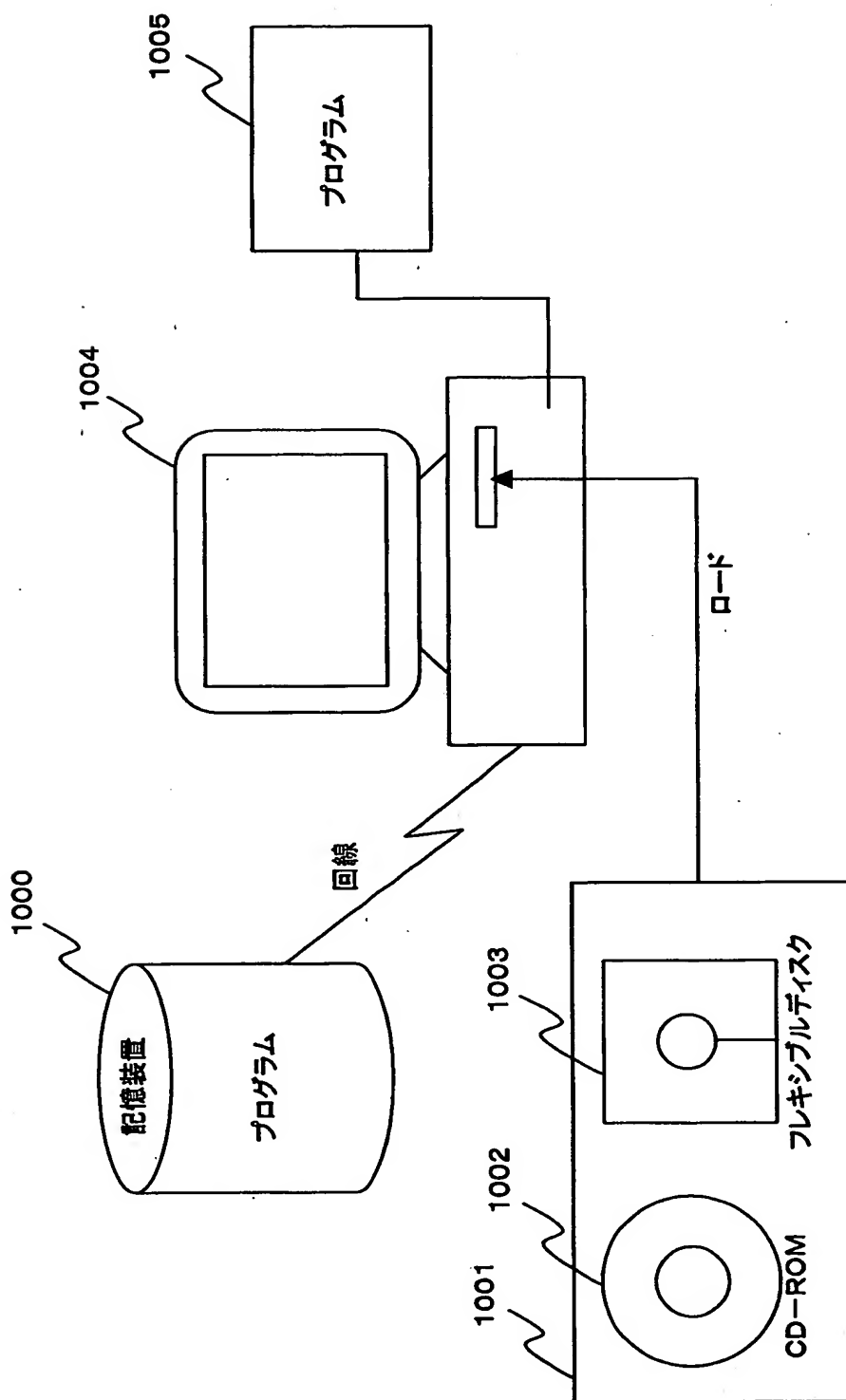


【図 15】





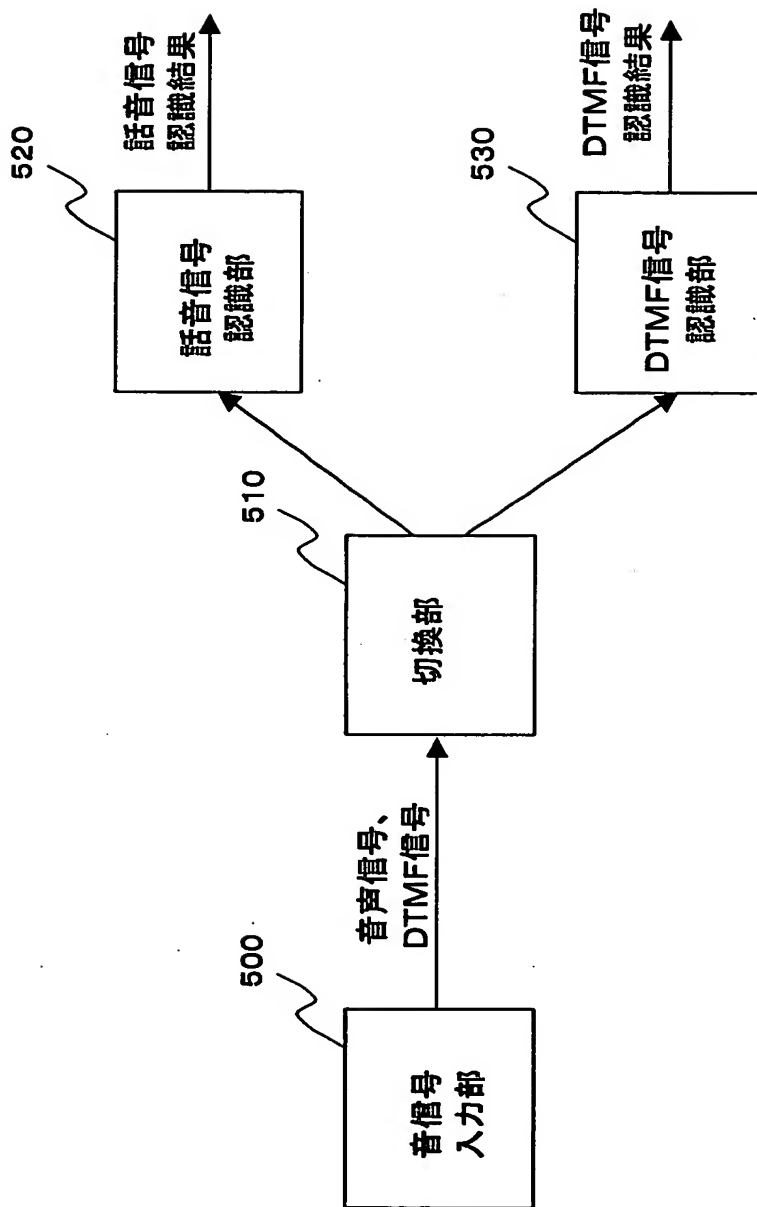
【図 16】



【図 1 7】

周波数[Hz]	1209	1336	1477	1633
697	1	2	3	A
770	4	5	6	B
852	7	8	9	C
941	*	0	#	D

【図 1 8】



【書類名】 要約書

【要約】

【課題】 音信号入力が、話音信号のみ、DTMF信号のみ、話音信号区間およびDTMF信号区間の双方が混在した音信号のいずれであっても、正しく認識し、入力モードの切り換え操作を不要とする音信号認識システムを提供する。

【解決手段】 話音信号区間またはDTMF信号区間のいずれか一方または双方を含む音信号を音信号入力部100を介して音信号照合・認識部200に入力する。音信号を音信号分析部210で音信号区間に分ける。照合部240は、DTMF信号モデル220および話音信号モデル230の双方を参照して音信号の照合処理を行ない、認識部260は単語辞書および文法規則情報を含む言語モデル250を備え、照合部240の照合結果を基に言語モデル250を用いて音信号の認識を行なう。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 5 2 2 3 ]

1. 変更年月日 1 9 9 6 年 3 月 2 6 日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号

氏 名 富士通株式会社